FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Football Transfers Network Analysis

ProDEI040 - Analysis of Social and Information Networks
2017/2018

Alexandre Ribeiro
up201205024@fe.up.pt

Introduction

- This project aims to study **football transfers**, with focus on their **values**, from a **social network perspective**
- **Gephi** was used as graph analysis tool
- Original data was scraped from https://www.transfermarkt.com/
  - English, Spanish, German, Italian, French and Portuguese leagues
  - 2015/2016, 2016/2017 and 2017/2018 seasons
  - Some transfers were **ignored**
    - Transfers from or to non main teams (i.e. U21, B, ...)
    - Transfers with unknown value, zero cost and loans

Dataset

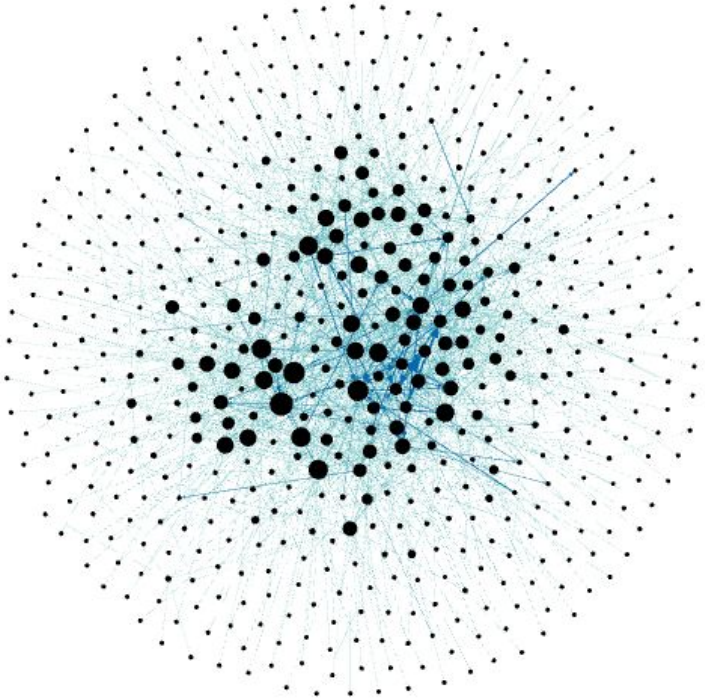
- The dataset consists of a **list of transfers**
  - Each **transfer** is a network **edge**
    - Player Name, Source Club, Target Club, Transfer Value
      - Transfer value, in euros, is the edge **weight**
  - Each **club** is a network **node**
  - There are **2151** transfers

Player	Source	Target	Value
Corentin Tolisso	Olympique Lyon	Bayern Munich	41500000

Chapter 1 and 2

Overview and Graph

The Graph - First Overview



Context ×

Nodes: 557

Edges: 1953

Directed Graph

- Fruchterman Reingold Layout
- Less edges than dataset entries
  - Edges with same (Source Club, Target Club) pair were **merged** and their **weights added together**

Some measures of the network

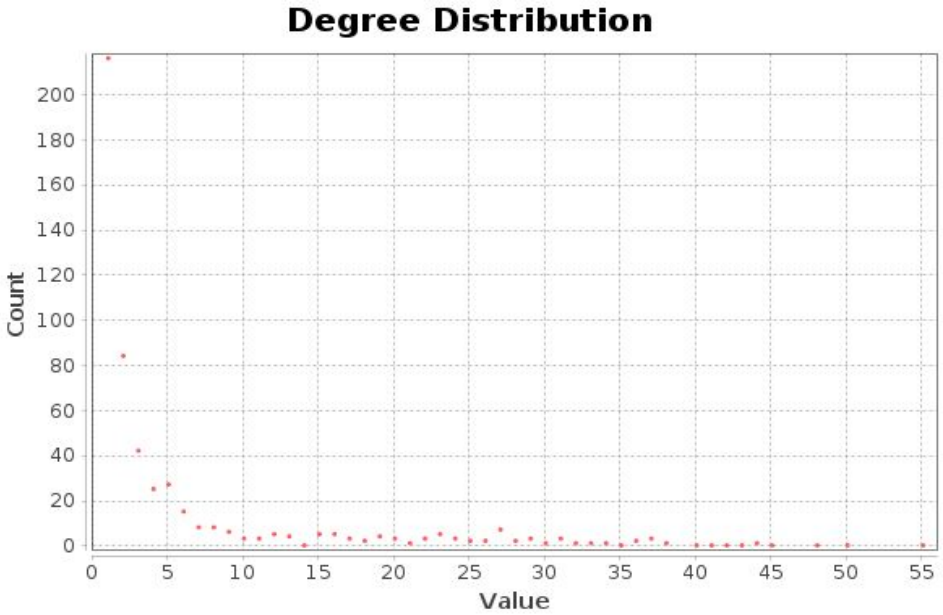
- Average Node Degree
  - 3,506
- Average Weighted Degree
  - 25669685

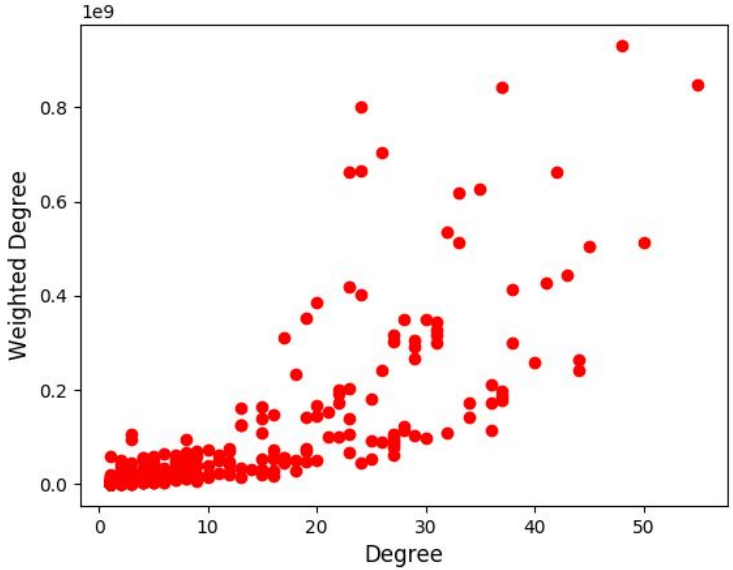
→

- In average, each team was involved in 3,5 transfers
- In average, each team moved almost 26 million euros

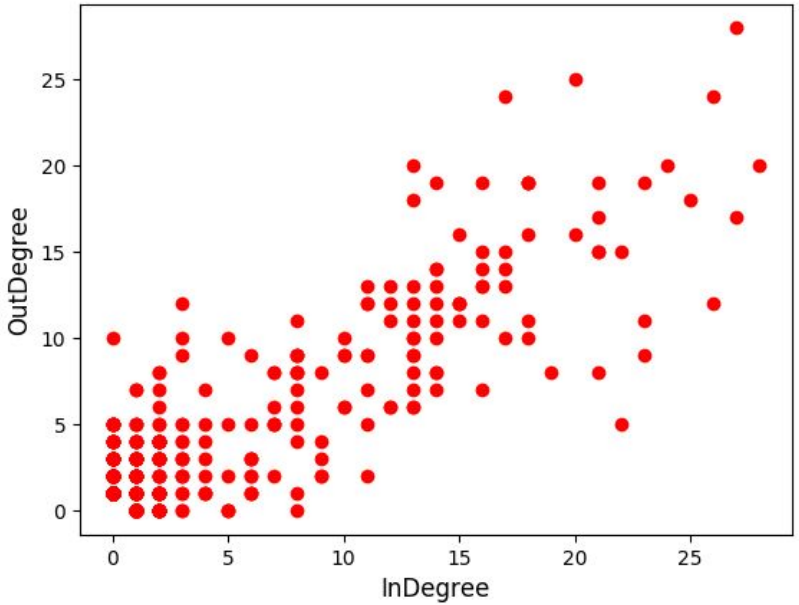
	Average Path Length	Network Diameter
Directed	3,599	7
Undirected	3,339	6

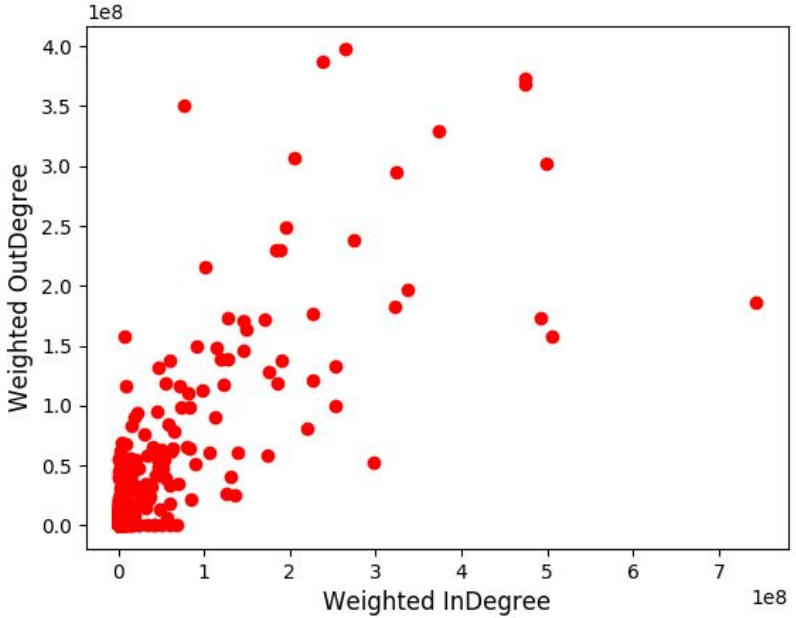
Some measures of the network (cont.)





Some measures of the network (cont.)

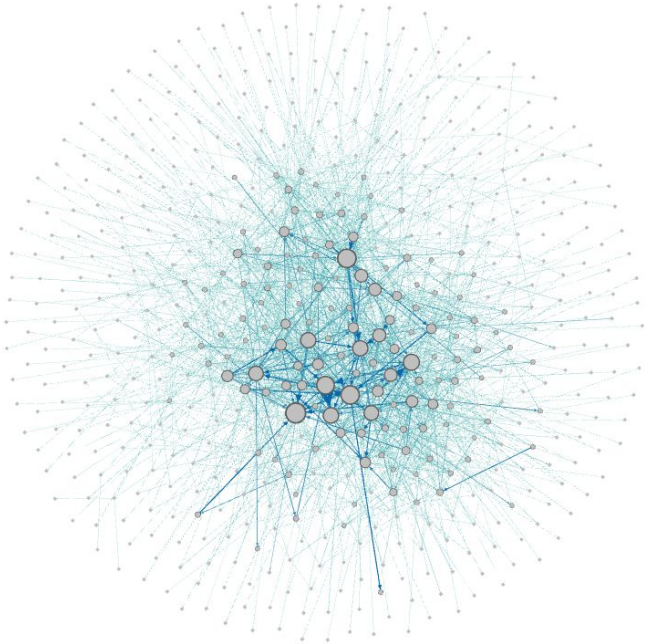




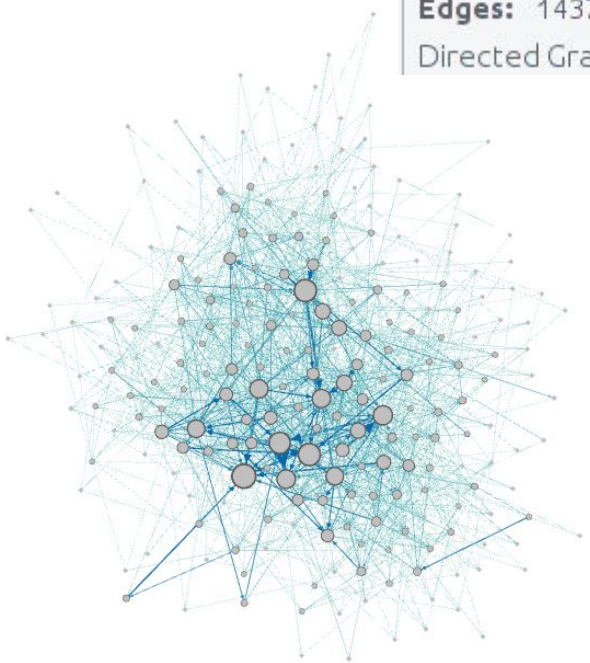
Connected Components

- This football transfers dataset comprises:
  - 1 weakly connected component
  - 327 strongly connected components
- A directed graph is **strongly connected** if there is a **path in each direction** between **each pair of vertices**
- The **giant component** is useful to depict the **core routes** of football players

Strongly Connected Components



a) Entire Network



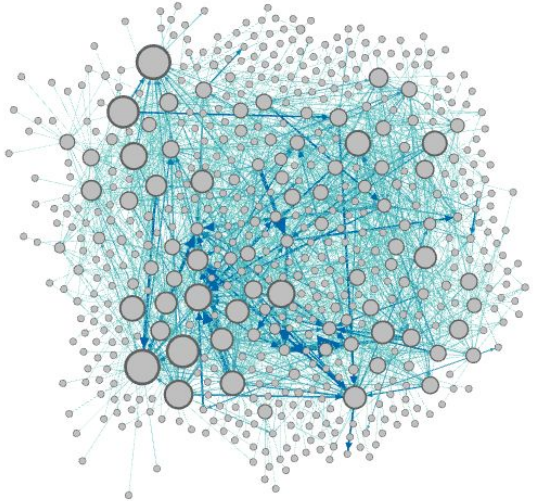
b) Network Giant Component

Chapter 3

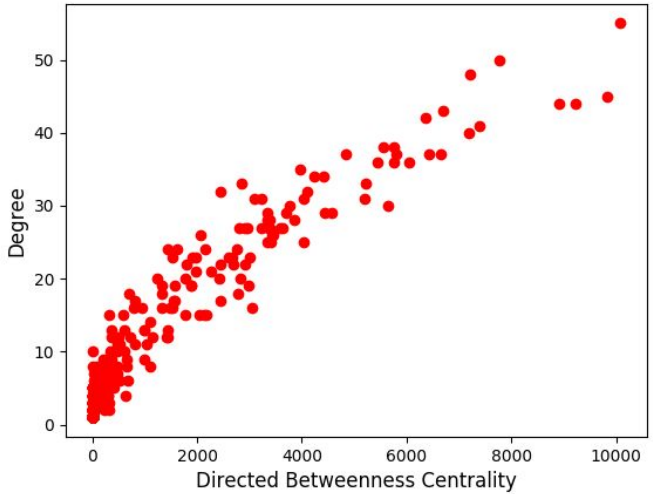
Strong and Weak Ties

Directed Betweenness Centrality

- Clubs acting as transfer brokers
- Clubs with ability to connect other clubs



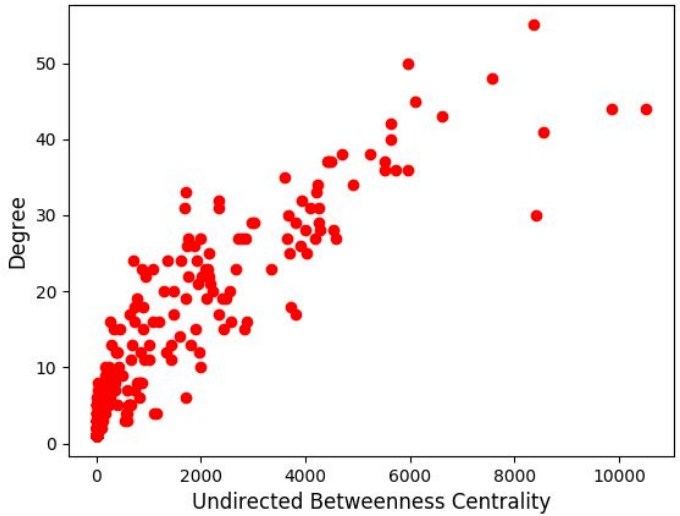
Id	Betweenness Cen...
Juventus FC	10080.762513
Inter Milan	9827.375422
ACF Fiorentina	9214.215949
Sporting CP	8911.397713
AS Roma	7776.889477
SL Benfica	7398.874308
Manchester City	7214.960602
UC Sampdoria	7191.315191
VfL Wolfsburg	6686.104189
Torino FC	6656.99053



Nodes sized by betweenness centrality

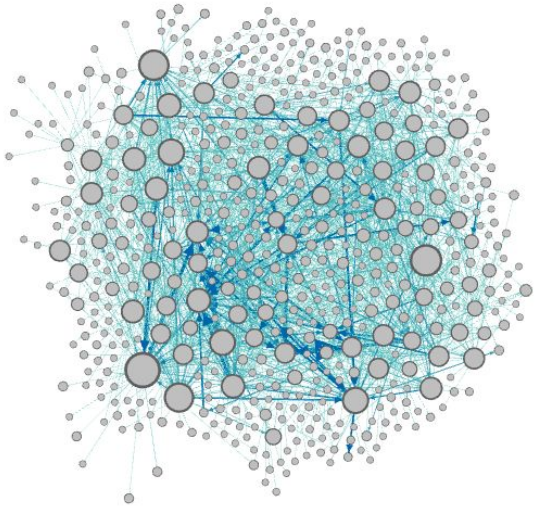
Undirected Betweenness Centrality

Id	Betweenness Cen...
ACF Fiorentina	10526.085479
Sporting CP	9857.706696
SL Benfica	8552.878445
SC Braga	8410.703103
Juventus FC	8375.924342
Manchester City	7565.648989
VfL Wolfsburg	6618.559115
Inter Milan	6092.497965
AS Roma	5973.893891
LOSC Lille	5951.935111



Eigenvector Centrality

- Represents the relative influence of a club in the network
- Based on the density of the inbound connections (purchases) to each club and its neighbors



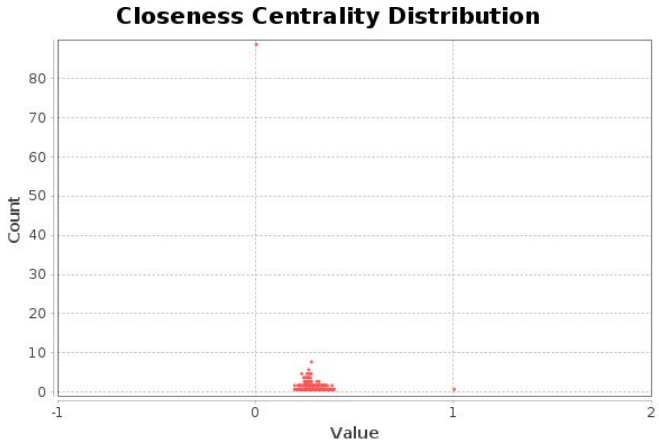
Nodes sized by eigenvector centrality

Id	Eigenvector Centrality
Juventus FC	1.0
Watford FC	0.86546
Inter Milan	0.831948
AS Roma	0.79393
AC Milan	0.699482
Chelsea FC	0.682429
Sevilla FC	0.675637
Atalanta BC	0.611128
Manchester City	0.609273
AS Monaco	0.593155

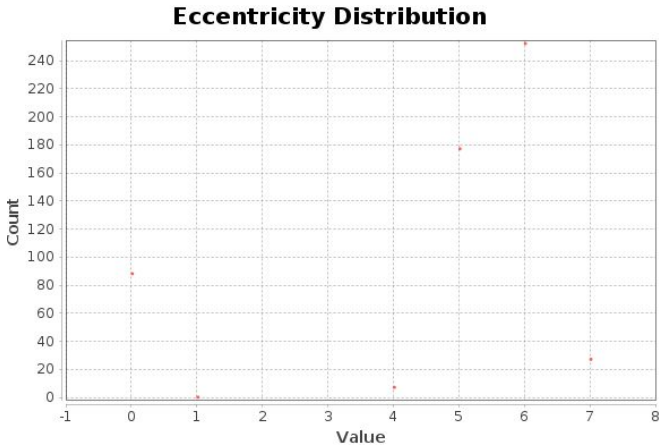
Larger means stronger attraction to players

Closeness Centrality & Eccentricity

- Average distance from a node to all other nodes in the network
  - Lower values are mainly clubs outside the selected leagues

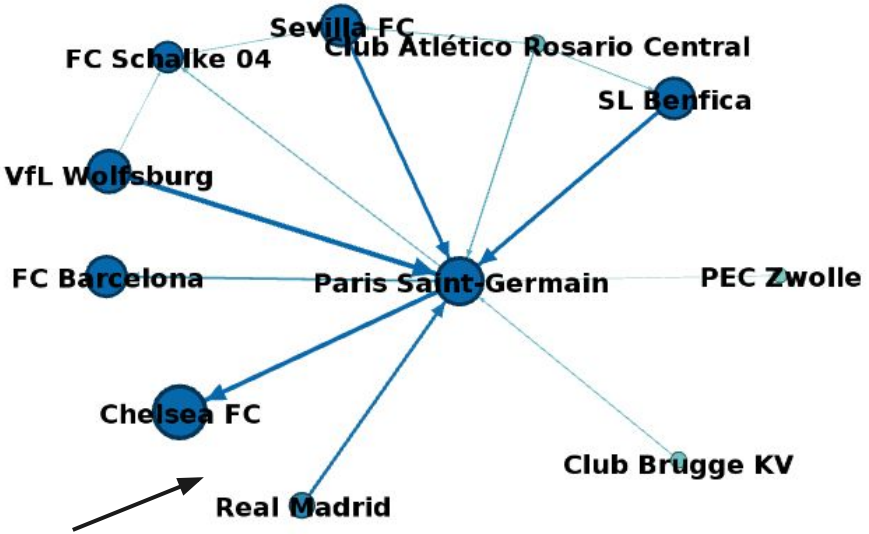


- Distance from a node to the farthest node from it in the network
  - Ranges from zero to network diameter



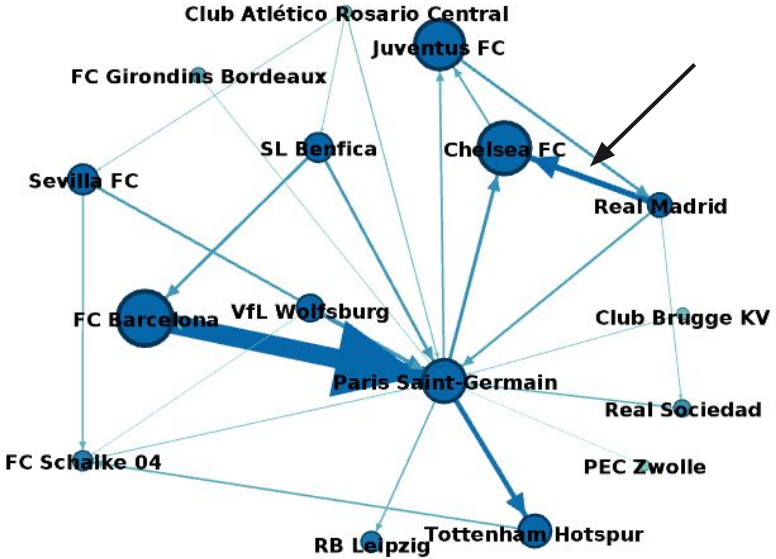
Triadic Closure

- Season 2016
- There is a strong tie between Paris Saint-Germain and Chelsea FC and Real Madrid



Triadic Closure

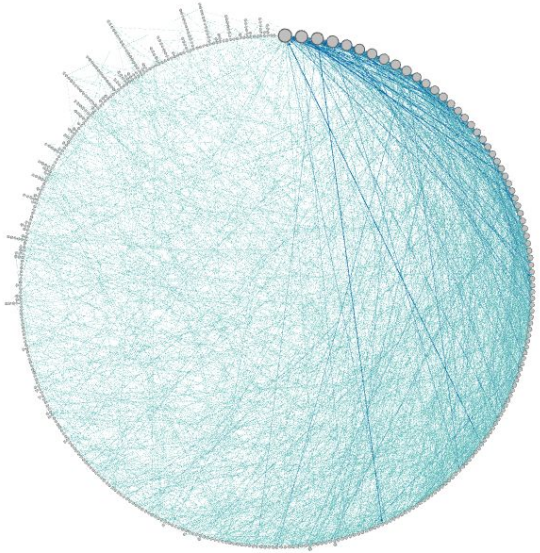
- Season 2017
- A new tie is formed between Chelsea FC and Real Madrid



Chapter 4

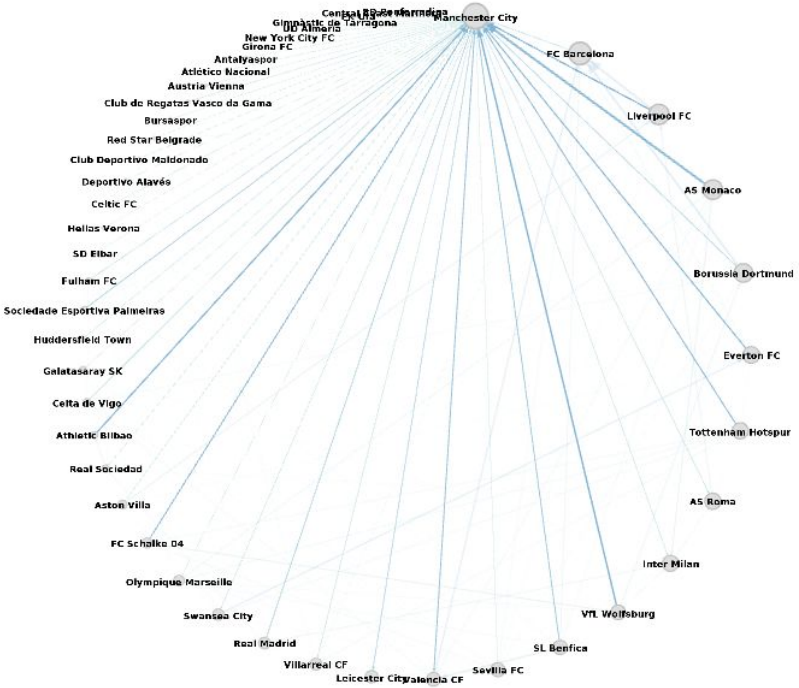
Networks in Their Surrounding Contexts

Homophily

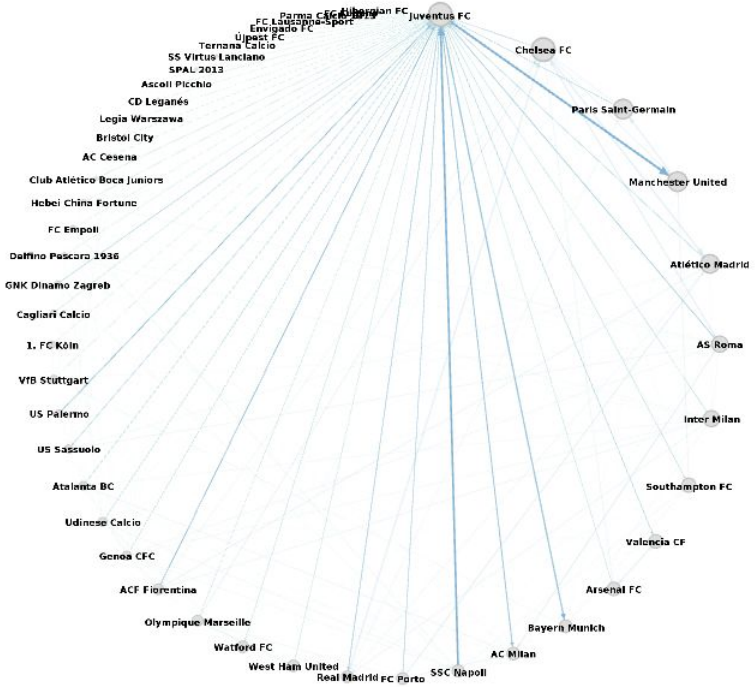


- Similar nodes may be more likely to attach to each other than dissimilar ones
- Assuming the **weighted degree** as the node similarity measure
- Nodes **grouped** by weighted degree

Homophily

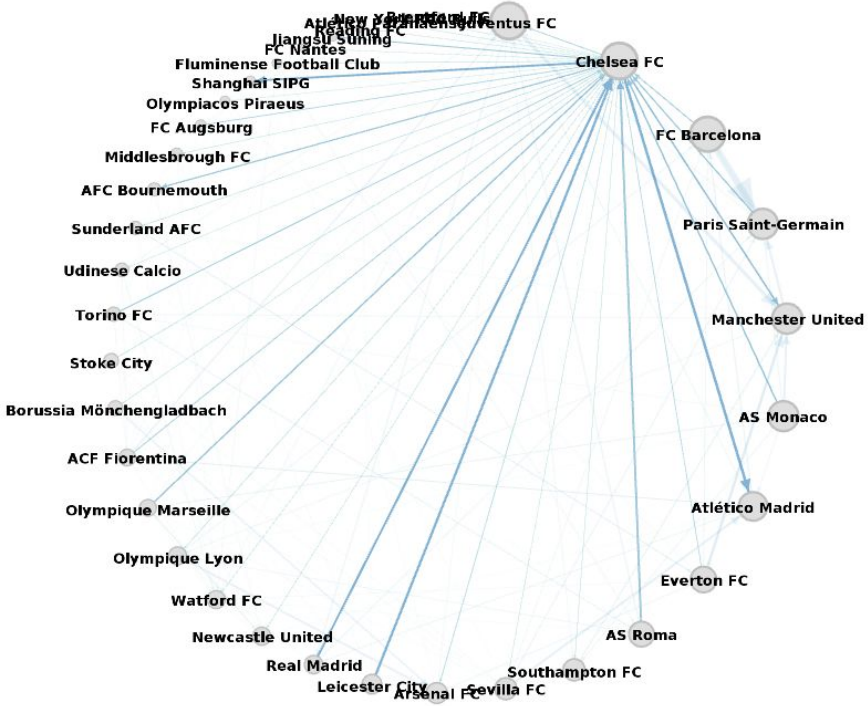


Manchester City Ego-Network

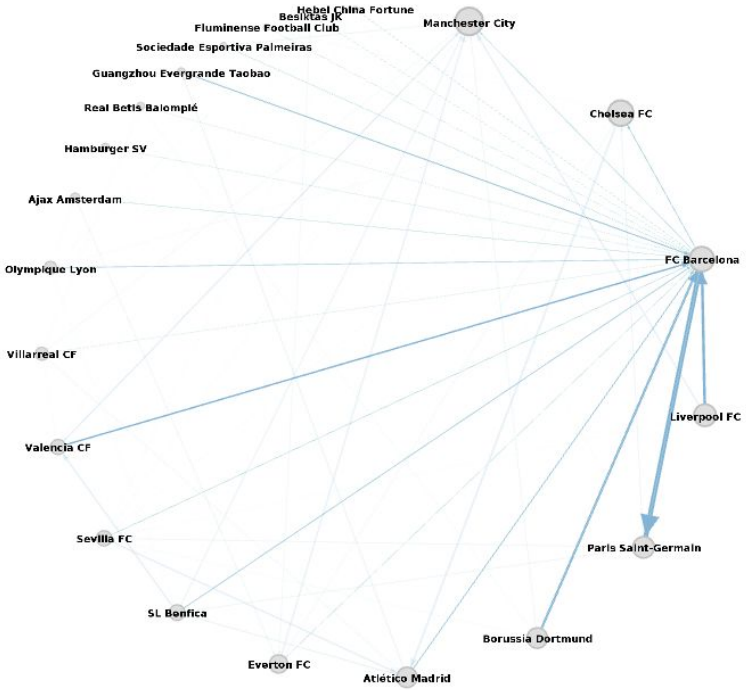


Juventus FC Ego-Network

Homophily



Chelsea FC Ego-Network



FC Barcelona Ego-Network

Homophily

- The homophily principle is **not rigorously verified**
- Generally, the clubs relate **without major restrictions**
    - They seek for players who meet their needs and budget
    - If they are direct competitors, they may not easily transfer between them

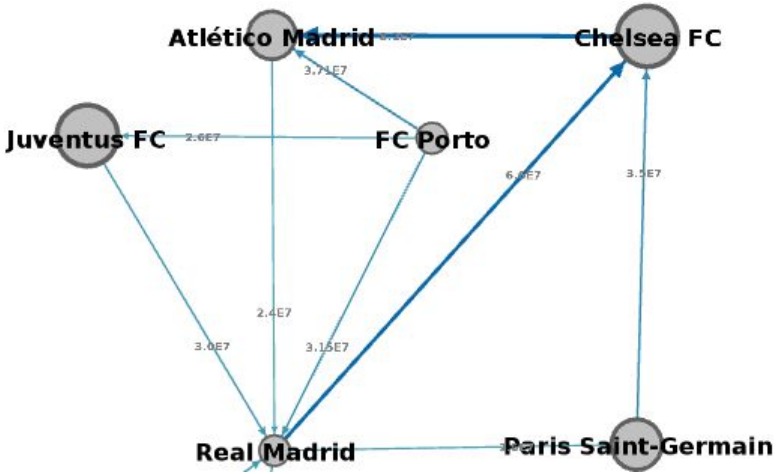
Chapter 5

Positive and Negative Relationships

Structural Balance

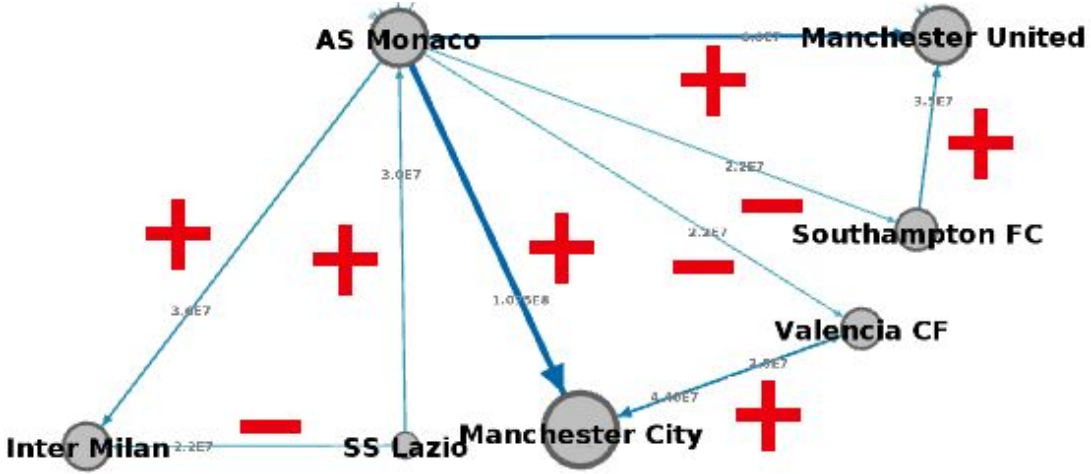
- For every set of **three nodes**, if we consider the three edges connecting them, **either all three** of these edges are labeled positive, **or else exactly one** of them is labeled positive

Positive if clubs trade **20 million** euros or more



Balanced: all edges are positive

Positive if clubs trade **25 million** euros or more

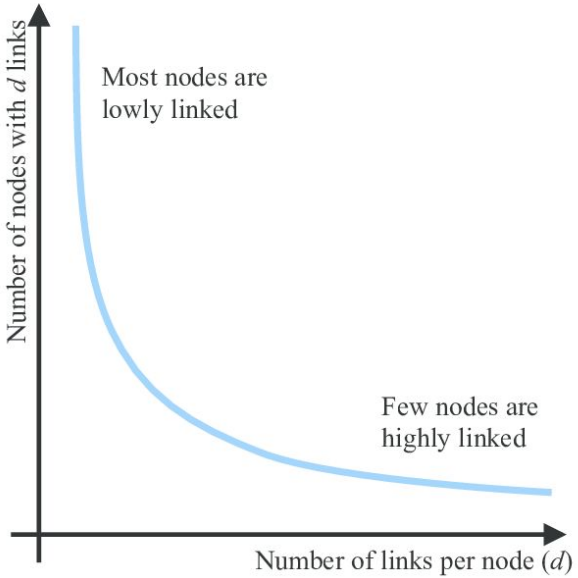


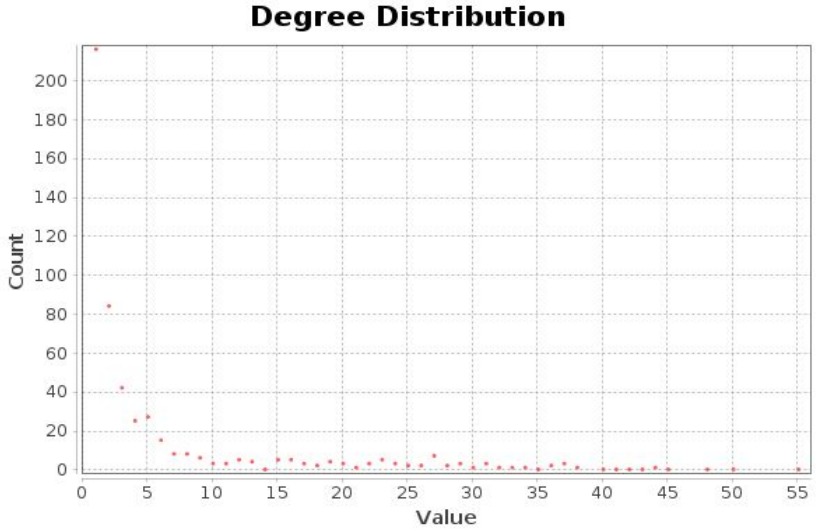
Not Balanced: each triangle has two positive edges

Chapter 18

Power Laws and Rich-Get-Richer Phenomena

Power Law





Example of power law distribution

Rich-get-Richer Phenomena

- This phenomenon is also known as **preferential attachment**
  - Links are formed “preferentially” to clubs that already have high degree
- Wolverhampton Wanderers **doesn’t exist** in the **2016** dataset
  - If we take into account the **2016 + 2017** dataset
    - The **average degree** is 2,869
    - It mainly linked with clubs with high degree

Id	Degree
SL Benfica	31
AS Monaco	24
FC Porto	18
SCO Angers	15
Rio Ave FC	6

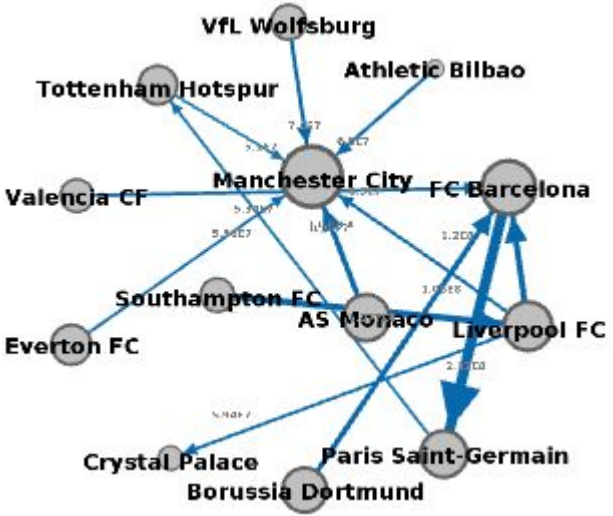
Chapter 19

Cascading Behavior in Networks

Cascading

Considering a **high value transfer (50 million euros or more)** as a behavior that may be propagated

2016 Liverpool FC Ego-Network **Depth 1**



2016+2017 Liverpool FC Ego-Network **Depth 2**

Chapter 20

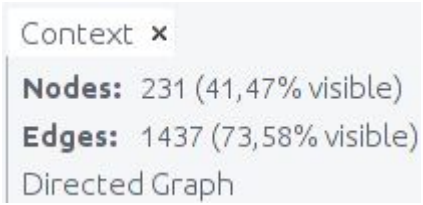
The Small-World Phenomenon

The Small-World Phenomenon

- Considering the network as **undirected**, it **exhibits** the small-world phenomenon
  - Average shortest path length: 3.339
  - Average clustering coefficient: 0.176
  - Diameter: 6
- If we consider it as **directed** the small-world phenomenon **doesn't hold**
  - There are clubs without transfers-in or transfers-out, making them points without return

The Small-World Phenomenon

- Considering the **biggest strongly connected component,** it **exhibits** the small-world phenomenon, despite **weaker** than the undirected network
  - Average shortest path length: 3.226
  - Average clustering coefficient: 0.116
  - Diameter: 7



Some interesting clubs

- **Juventus**
  - More transfers, in and out (highest degree)
  - Most central club (highest betweenness and eigenvector centrality)
- **Manchester City**
  - Moved more money (highest weighted degree)
    - Club that spent more in transfers in (highest weighted in-degree)
    - Biggest difference between weighted in-degree and weighted out-degree
- **Leeds United**
  - Highest degree between clubs with no triangles in their neighbour

Conclusions

- Clubs relate **without major restrictions**
- Italian clubs are the most active
- **More context** would be interesting
    - **Timestamps** of transfers
    - Relate the **money spent** by a team with its **performance** (UEFA ranking)